DOCUMENT RESUME

ED 068 525 TM 001 886

AUTHOR	Loadman, William E.
TITLE	An Inquiry Concerning the Use of Item Sampling as a Method to Reduce Testing Time.
INSTITUTION	Indiana Univ., Bloomington.
REPORT NO	EREL-RP-5
PUB DATE	72
NOTE	14p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Comparative Analysis; Data Collection; Educational Innovation; Elementary Education; Elementary Grades; *Item Sampling; *Measurement Techniques; *Questionnaires; *Research Methodology; Role Perception; Statistical Data; *Time

ABSTRACT

The procedure of item sampling was employed to reduce the time expenditure of participants when responding to a questionnaire concerned with the implementation of an innovative elementary school project. Approximately 50 student teachers and 42 regular classroom teachers responded to one of two forms of the questionnaire. There was a total of 103 multiple choice questions on each form with 13 questions being common to both forms (a total of 193 different questions). Participation time was reduced from two hours to one hour for each participant. A comparison between the perception of student teachers on form A and B using the common items showed no significant differences; the same result held for the classroom teachers. However, there were differences between the perceptions of student teachers and the classroom teachers. These data were used to estimate and compare the perceptions of both groups for all items for all participants. (Author)

ED 068525

RESEARCH PAPERS

Educational Research and Evaluation Laboratory

INDIANA UNIVERSITY

Tid 001 886

Not to be cited
without permission

ED 068525

AN INQUIRY CONCERNING THE USE OF ITEM SAMPLING AS A METHOD TO REDUCE TESTING TIME

William E. Loadman
Indiana University
1972

No. 5

AN INQUIRY CONCERNING THE USE OF ITEM SAMPLING AS A METHOD TO REDUCE TESTING TIME

William E. Loadman
Indiana University

Abstract

This study employed the procedure of item sampling to reduce the time expenditure of participants when responding to a questionnaire concerned with the implementation of an innovative elementary school project. Approximately 50 student teachers and 42 regular classroom teachers responded to one of two forms of the questionnaire. There was a total of 103 multiple choice questions on each form with 13 questions being common to both forms (a total of 193 different questions). Participation time was reduced from two hours to one hour for each participant. A comparison between the perception of student teachers on form A and B using the common items showed no significant differences; the same result held for the classroom teachers. However, there were differences between the perceptions of student teachers and the classroom teachers. These data were used to estimate and compare the perceptions of both groups for all items for all participants.

One of the problems evaluators continually face is gaining the cooperation of program participants so that reasonable evaluation efforts may be initiated. The time committment required to complete questionnaires, fill out information sheets, and comply with the wishes of the evaluator (among a host of others) are typical reasons given for the less than ideal cooperation received by the evaluator. Let it suffice to say that the priorities of the evaluator and the priorities of program participants are not always consonant.

Recently an effort was made to alleviate some of this disonnance by reducing the amount of time required by any program participant in completing questionnaire data. There are obvious ways to eliminate large amounts of time from evaluative sessions such as: (a) keeping the number of sessions to a minimum; (b) making the task of the respondee easily understandable, simple and straight forward; (c) keeping the number of open ended questions to a minimum and when using open ended questions, structure the question so that the answer will be brief and to the point; (d) trying to limit the number of items; (e) reducing the duplication of information, e.g., by not asking for a person's school, age, grade, etc. at each session; and (f) making the items unambiguous, i.e., keep the vocabulary simple and use only one thought or concept per item. Another way to reduce time committments is through the use of item sampling. This procedure has either not been obvious or evaluators have been avoiding its use.

The purpose of this study was to determine the feasibility of using an item sampling procedure to estimate the perceptions of a large group pf persons with only half of the group responding to any one question. A second purpose was to try to reduce the amount of time a given project participant would have to commit to completing an evaluative questionnaire. According to Lord (1962) item sampling procedures were found to be appropriate for estimating test performance of a group of individuals. This study attempted to extend the techniques for estimating group performance on questionnaire data. Since individual results were of no particular interest in this evaluation, the method of item sampling was employed.

REVIEW OF RELATED LITERATURE

There has been increasing interest in the use of item sampling since Lord (1962) published an article on the topic. The primary emphasis of the technique has been associated with achievement testing, i.e., using samples of items and/or samples of individuals to estimate group achievement on all items or all individuals. The mathematical formulation of the procedure is documented in Lord and Novick (1968). Shoemaker (1971) presents a lucid description and application of the technique along with the appropriate formulas and recommendations for use of the procedure. Several other studies dealing with item sampling (matrix sampling) have recently appeared in the literature (Cook and Stuffle beam, 1967; Johnson and Lord, 1958; Plumlee, 1964; and Shoemaker, 1970a, 1970b). These works have been primarily

concerned with emperical investigations of the validity of the model as applied to achievement test data. The results of these studies revealed that the item samples could be used to accurately predict group achievement on the entire assessment measure without all subjects responding to all items.

Sirotnik (1970) investigated the effect of different item contexts on subjects responses. These results along with similar studies by Shoemaker (1970c) and Burton and Remer (1972) indicated that there were minimal contextual effects associated with this procedure. The work of Burton and Remer dealt with contextual effect using questionnaire data. Pugh (1971) investigated the use of item sampling procedures with Likert scale items and found the procedure to yield accurate estimates of central tendency and variability.

An immediate extension of the technique would suggest its potential use with questionnaire data. This study was an initial step in ascertaining the appropriateness and feasibility of the technique with questionnaire data.

PROCEDURE

Fifty student teachers and 42 elementary classroom teachers associated with a large midwestern university were asked to respond to a questionnaire concerning an innovative elementary school program. Because there was a large number of items (193), the researcher decided to build two forms of the questionnaire. One form of the

questionnaire contained a random sample of one half of the total number of items and the second form of the questionnaire contained the remainder of items. In addition, thirteen of the items were purposely placed on both forms to allow for esitmates of group comparability. Thus there was a total of 103 items on each form.

The two forms of the questionnaire were randomly assigned to program participants within each group (student teachers and classroom teachers). Twenty-six and 24 student teachers responded to Form A and Form B respectively. Twenty-two and 20 classroom teachers responded to Form A and Form B respectively. The items on each questionnaire were concerned with the implementation and value of many components of the innovative project. Each item contained a five point response continuum upon which the respondee indicated his estimated agreement, value, worth, etc.

The responses were coded, placed on computer cards and a preliminary analysis of the 13 common items was undertaken. A 4 x 13 repeated measures analysis of variance was initiated. However, a significant groups by measures interaction necessitated additional analyses. Thirteen one way analyses of variance were conducted using the level of significance of $\alpha < .10$. This liberal alpha level was used so as to not miss significant differences among the groups. Following the preliminary analyses, a binomial test was applied to the thirteen ANOVA outcomes for each of the groups. The

purpose of this test was to ascertain the exact chance probability of the thirteen outcomes.

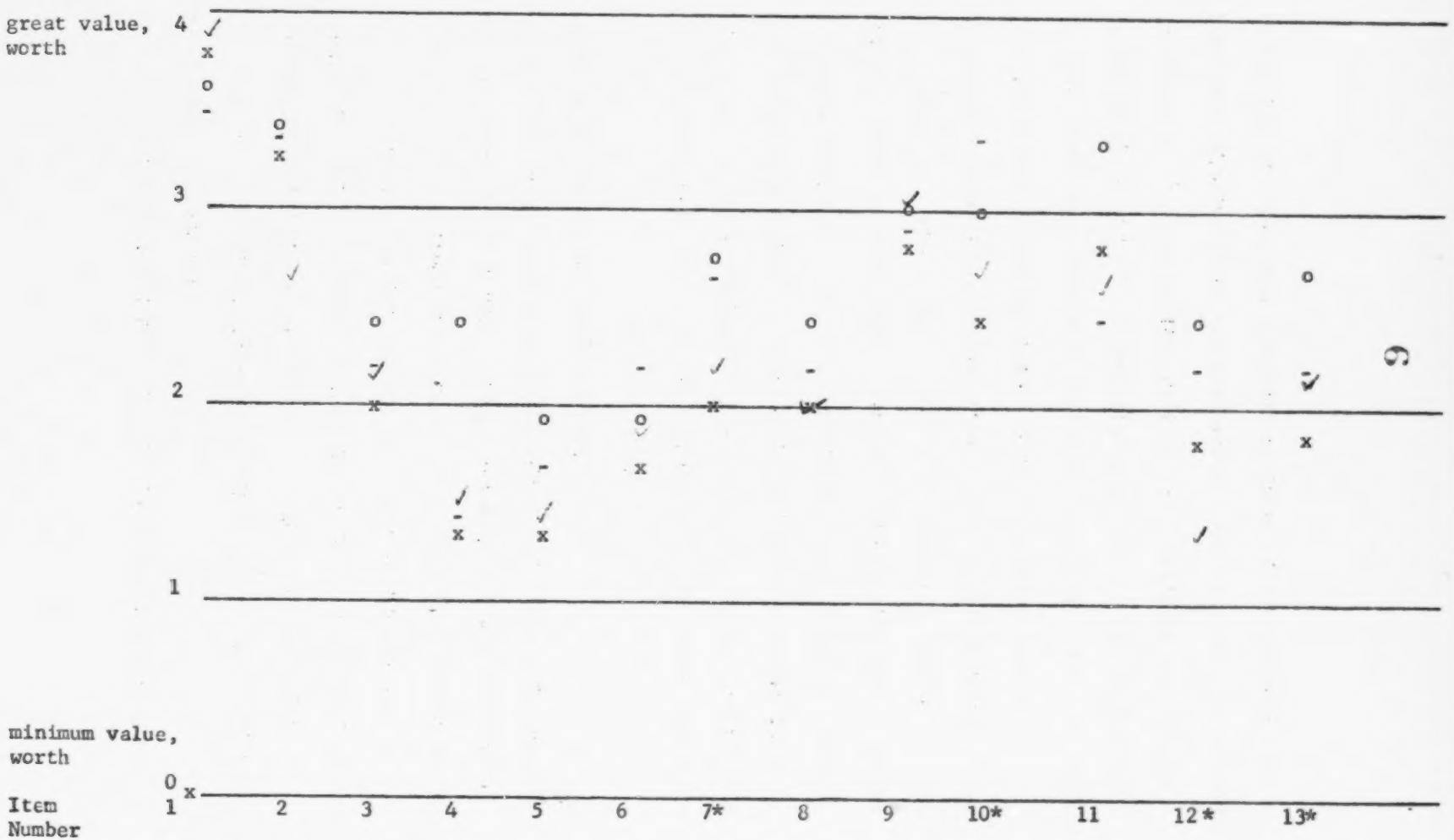
Using the results of the ANOVA's and the binomial tests, a decision was made to use the sample data from one group (e.g., student teachers responding to Form A) to estimate the performance of the combined group (all student teachers within the program). Since none of the items were placed in clusters, but rather analyzed item by item, the best estimate of the combined group score on a given item was the sample data from one group. This translated into measures of central tendency and variability.

RESULTS AND DISCUSSION

On the initial analyses, each of the thirteen items were submitted to a one-way analysis of variance. The student teacher groups were found to not have any significant differences on the 13 items. A similar result was obtained from the analysis of the classroom teacher data. However, significant differences were found between the responses of the classroom and student teacher groups. Significant differences between the student teacher and classroom teacher groups were found on 5 of the 13 items, $\alpha < .10$ (see Figure 1).

Following these analyses, a binomial test was run on the outcomes of the ANOVA's. Using a probability value of success of $p = .90$, the probability of obtaining thirteen nonsignificant differences among the thirteen items was determined. This probability was equal to .25 for the classroom teachers; obviously a similar value was

Figure I - Program Ratings - Group Means



obtained for the student teacher. The results of the binomial test can be interpreted to mean that if the null hypothesis were true (i.e., there are no true differences between the two groups of classroom teachers,) then the probability of finding thirteen non-significant differences on thirteen trials by setting $\alpha < .10$ is equal to .25.

A critical factor in the results and interpretation of an item sampling procedure must be within cell variability. It is apparent from figure 1 that the within cell variability across items must be different because some group means (absolute score differences) were not far apart and yet there was a statistically significant difference among group means (e.g., item 1). However, other items showed greater absolute differences across group means and yet these differences were not statistically significant (e.g., item 4). The within cell variances of the 13 items ranged from a low of .25 to a high of 2.8; the median within cell variance of these items was approximately 1.2.

On the basis of the above analyses, each of the remaining 90 items on the questionnaires was analyzed and used to estimate total group performance. This study has immediate ramification for generalizing the results of a small sample to the larger sample and for survey research in general.

Using as a guide the ANOVA model, the following rationale may be employed in analyzing the remaining data and estimating total group performance. It was assumed that the within group variability for

persons responding to an item was an accurate and unbiased estimate of the within group variability for the group not responding to the item. Total within group variability could then be estimated by pooling the separate within group variabilities. This procedure would allow for an estimate of total group variability, (equal to the variability within a subgroup) while increasing degrees of freedom. Therefore a statistical test of significance could be applied to the two complete groups of data.

Rather than using this liberal approach, a more conservative test was applied to the data. The analysis was conducted on the original data, using the reduced degrees of freedom and the unpooled within group variability. Significant differences between the classroom and student teaching groups were consistently found on this aspect of this analysis. The results of these analyses will not be discussed in this presentation.

In this study the time expenditure for each subject was approximately one hour. However, if each subject had responded to all 193 items it would have required a minimum of 2 hours of subject time. It was apparent that a fatique factor began to enter the picture as the subjects completed the one hour of information gathering, thus reinforcing this researcher's personal satisfaction with the sample procedure.

The results of this study suggested the feasibility and beneficiality of item and person sampling as an appropriate methodology to estimate

group performance when using questionnaire responses. This procedure has the obvious advantages of (a) reducing the time commitment of any single participant in completing questionnaires, (b) providing accurate estimates of total group performance based upon a small sample of data; and (c) allowing for a greater variety and number of questions to be asked.

A word of caution must be introduced with the use of this procedure. As most researchers are aware, the stability of measurement is reduced as the number of observations is decreased. When the number of observations is small (10 or less) the stability of the group estimates begin to fluctuate and may be highly susceptible to Type I and Type II errors. Also, the nonindependence among items may yield an uncommon number of Type I errors. Therefore, caution must be exercised in evaluating this and other similar data.

SUMMARY

This study tested the appropriateness of an item-person sampling procedure applied to questionnaire data. With these data (smallest subgroup = 20 observations) the model was found to be time saving, feasible and accurate. On thirteen items that were common to two forms of the questionnaire two subgroups of student teachers did not significantly disagree on any of the items. Two subgroups of classroom teachers responses yielded similar results. However, there were significant responses between the responses of classroom and student teachers.

The expectation that these results would occur by chance was found to be relatively small.

The respondees' time was significantly reduced with this procedure without reducing the generalizability of the results. Its ease of usage and apparent accuracy would suggest increased adoption of the technique. This procedure will probably also enhance the cooperation of the respondees. In addition, there seems to be great promise for the technique applied to questionnaire and attitude scale data.

BIBLIOGRAPHY

Burton, N. and Remer, Roy. Item Sampling in Questionnaires: Rate of Return and Context Effect. 1972 AREA Paper.

Cook, D. L. and Stufflebeam, D. L. Estimating test norms from variable size item and examinee samples. Educational and Psychological Measurement, 1967. 27, 601-610.

Johnson, M. C. and Lord, F. M. An emperical study of the stability of a group mean in relation to the distribution of test item among pupils. Educational and Psychological Measurement, 1958, 18, 32F329.

Lord, F. M. Estimating norms by item sampling. Educational and Psychological Measurement, 1962, 22, 259-267.

Lord, F. M. and Novick, M. R. Statistical theories of Mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Plumlee, L. B. Estimating means and standard deviations from partial data---an empirical check on Lord's item sampling technique. Educational and Psychological Measurement, 1964, 24, 623-630.

Pugh, R. C. "Emperical Evidence on the Application of Lord's sampling technique to Likert items." Journal of Experimental Education, 1971, 39, 54-56.

Siegel, S. Nonparametric Statistics for the behavioral sciences, New York: McGraw-Hill, 1956.

Sirotnik, K. An investigation of the context effect in matrix sampling. Journal of Educational Measurement, 1970, 7, 199-207.

Shoemaker, D. M. Allocation of item and examinees in estimating a norm distribution by item-sampling. Journal of Educational Measurement, 1970, 7, 123-128.

Shoemaker, D. M. Item-examiner sampling procedures and associated standard errors in estimating test parameter. Journal of Educational Measurement, 1970, 7, 255, 262.

Shoemaker, D. M. Test statistics as a function of item arrangement. Journal of Experimental Education, 1970, 39, 85-88.

Shoemaker, D. M. Principles and Procedures of Matrix Sampling. Southwest Regional Laboratory: Technical Report 34, 1971.